BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

GENOMICS 50, 357–367 (1998)
ARTICLE NO. GE985317

The CD39-like Gene Family: Identification of Three New Human Members (CD39L2, CD39L3, and CD39L4), Their Murine Homologues, and a Member of the Gene Family from *Drosophila melanogaster*

Brian Paul Chadwick and Anna-Maria Frischauf[1]

Molecular Analysis of Mammalian Mutation, Imperial Cancer Research Fund, London WC2A 3PX, United Kingdom

Received January 16, 1998; accepted March 23, 1998

The human lymphoid cell activation antigen *CD39* is a known E-type apyrase that hydrolyzes extracellular ATP and ADP, a function important in homotypic adhesion, platelet aggregation, and removal by activated lymphocytes of the lytic effect of ATP. The recently identified putative rat homologue of *CD39L1* has been shown to have E-type ecto-ATPase activity, by hydrolyzing extracellular ATP. We have characterized three novel CD39-like transcripts, CD39L2, CD39L3, and CD39L4, which share extensive amino acid homology with other nucleotide triphosphatases in vertebrates, invertebrates, and plants, suggesting that these genes also encode proteins with ecto-nucleotidase activity. Isolation and sequencing of full-length cDNA clones for each gene identified putative proteins of 485, 529, and 429 amino acids. The expression pattern of all five human members of the gene family was analyzed. *CD39L2*, *CD39L3*, and *CD39L4* were mapped on the human genome, and the murine homologues identified with the putative map locations were assigned on the basis of regions of conserved gene order between human and mouse chromosomes. The map location of *mcd39l4* places the gene within a region associated with audiogenic seizure susceptibility in mouse. This disorder is characterized by convulsions induced by loud high-frequency sound and has been shown to be associated with increased nucleotide triphosphatase activity.

INTRODUCTION

Extracellular nucleotides have been implicated as signaling molecules in many extracellular activities, including neurotransmission, cardiac function, platelet aggregation, muscle contraction and relaxation, vascular tone, secretion of hormones, immune responses, and cell growth (for reviews, see Chen *et al.*, 1995, and Zimmermann, 1994). The catabolism of extracellular nucleotides is mediated by several types of ecto-nucleotidases, including the E-type nucleotidases (Zimmermann, 1996).

E-type activity is defined as divalent cation dependent and is insensitive to inhibitors of P-type, F-type, and V-type ATPases (Plesner, 1995). E-type NTPases include ecto-ATPases and ecto-apyrases, which differ in the rate of nucleoside diphosphate (NDP) hydrolysis compared with the rate of ATP hydrolysis (ecto-apyrases show a significant rate of ADP hydrolysis when compared to ATP hydrolysis, while ecto-ATPases hydrolyze NDPs at ~1–2% of the rate of ATP hydrolysis).

Human CD39 protein (cluster of differentiation-39) is a transmembrane glycoprotein located on the surface of activated lymphocytes and displays ecto-apyrase activity (Wang and Guidotti, 1996). Several potential functional roles for CD39 have been suggested, including protection of lymphocytes from the lytic effect of extracellular ATP released by target cells (Filippini *et al.*, 1990; Di Virgilio, 1995) and involvement in the regulation of homotypic adhesion of activated B lymphocytes by a non-integrin-mediated pathway (Kansas *et al.*, 1991). More recently CD39 has been shown to be a critical component in thromboregulation, since it controls the levels of the platelet agonist, ADP (Marcus *et al.*, 1997). Other potential functions that have been proposed for ecto-NTPases are adenosine recycling (Plesner, 1995), regulation of ecto-kinase substrate concentration, demyelination and remyelination of axons (Felts and Smith, 1996), termination of purinergic signaling (including neurotransmission by ATP), and the development and maintenance of neurons (Zimmermann, 1996).

CD39 shares considerable amino acid homology with several nucleotidases including a chicken muscle ecto-ATPase (Kirley, 1997), a yeast guanosine diphosphatase (Abeijon *et al.*, 1993), an ATP-diphosphohydrolase (apyrase) from potato tubers (Handa and Guidotti, 1996), a garden pea nucleotide triphosphatase (Lin, 1989), and several NTPases from *Toxoplasma gondii* (Bermudes *et al.*, 1994). Four highly conserved regions of homology exist in the N-terminal region of each protein, termed the apyrase conserved regions I–IV

[1] To whom correspondence should be addressed. Telephone: (0171) 269 2319. Fax: (0171) 269 3479. E-mail: frischau@icrf.icnet.uk.

0888-7543/98 $25.00

(ACRs). Proteins containing all four of the ACRs are likely to display NTPase activity, and therefore the ACRs provide a useful indication of potential nucleotidase activity (Handa and Guidotti, 1996).

While previously ectoapyrases and ecto-ATPases were mainly isolated by functional assays at the protein level, homologies at the nucleic acid level have more recently led to the identification of several new members of the *CD39*-related family (Chadwick and Frischauf, 1997; Kegel *et al.*, 1997; Chadwick *et al.*, 1998) The human *CD39-like-1* gene (Chadwick and Frischauf, 1997) shows greater than 40% amino acid identity to the CD39 protein and greater than 60% identity to the chicken muscle ecto-ATPase (Kirley, 1997). The recently identified rat brain ecto-ATPase shows greater than 85% amino acid identity to CD39L1 and is likely to be the rodent homologue, therefore suggesting that CD39L1 is also an ecto-ATPase (Kegel *et al.*, 1997). Since database searches indicated the presence of further *CD39*-like genes, we screened for all likely candidates and cloned and further characterized them. Here we report the cloning, mapping, and tissue distribution of three new human members of the *CD39-like* gene family. We also identified the murine homologues of the human genes and the first member of the gene family from *Drosophila melanogaster*.

MATERIALS AND METHODS

Identification, isolation, and sequencing of cDNA clones for CD39L2, CD39L3, and CD39L4. The nucleotide sequence of *CD39* (Accession No. S73813), *CD39L1* (Accession No. U91510), and *mNTPase* (Accession No. AF006482) were used in TBLASTX searches against entries in the expressed sequence tag (EST) database at EMBL/GenBank, using the Bork server (http://www.bork.embl-heidelberg.de/). cDNA clones for highly homologous IMAGE EST entries were obtained from the Human Genome Mapping Project Resource Centre (HGMP, Hinxton, UK). DNA was prepared with QiaTip-100 (Qiagen), and the cDNA was sequenced by primer walking with a fluorescence labeled dye-terminator cycle sequencing kit according to the manufacturer's instructions (PRISM Ready Dye-Deoxy Terminator Premix from Applied Biosystems Inc) and electrophoresed on an ABI 373 (Perkin-Elmer). Overlapping EST clones were identified by searching with the nucleotide sequence against entries in the EST database using BLAST-N (http://www.ncbi.nlm.nih.gov:80/cgi-bin/BLAST/nph-blast?Jform).

Additional IMAGE cDNA clones were ordered from HGMP if they extended the existing nucleotide sequence further 5′. cDNA clones corresponding to the most 5′ extreme of each gene were identified by hybridization of radiolabeled inserts of IMAGE cDNA clones to a keratinocyte stem cell cDNA library (Phil Jones and Fiona Watt, ICRF), a human adult breast epithelial cDNA library constructed using Stratagene Lambda ZAP vector (Rodger White and Malcolm Parker, ICRF), and a Jurkat cell line cDNA library in pBluescript (Dunne *et al.*, 1995).

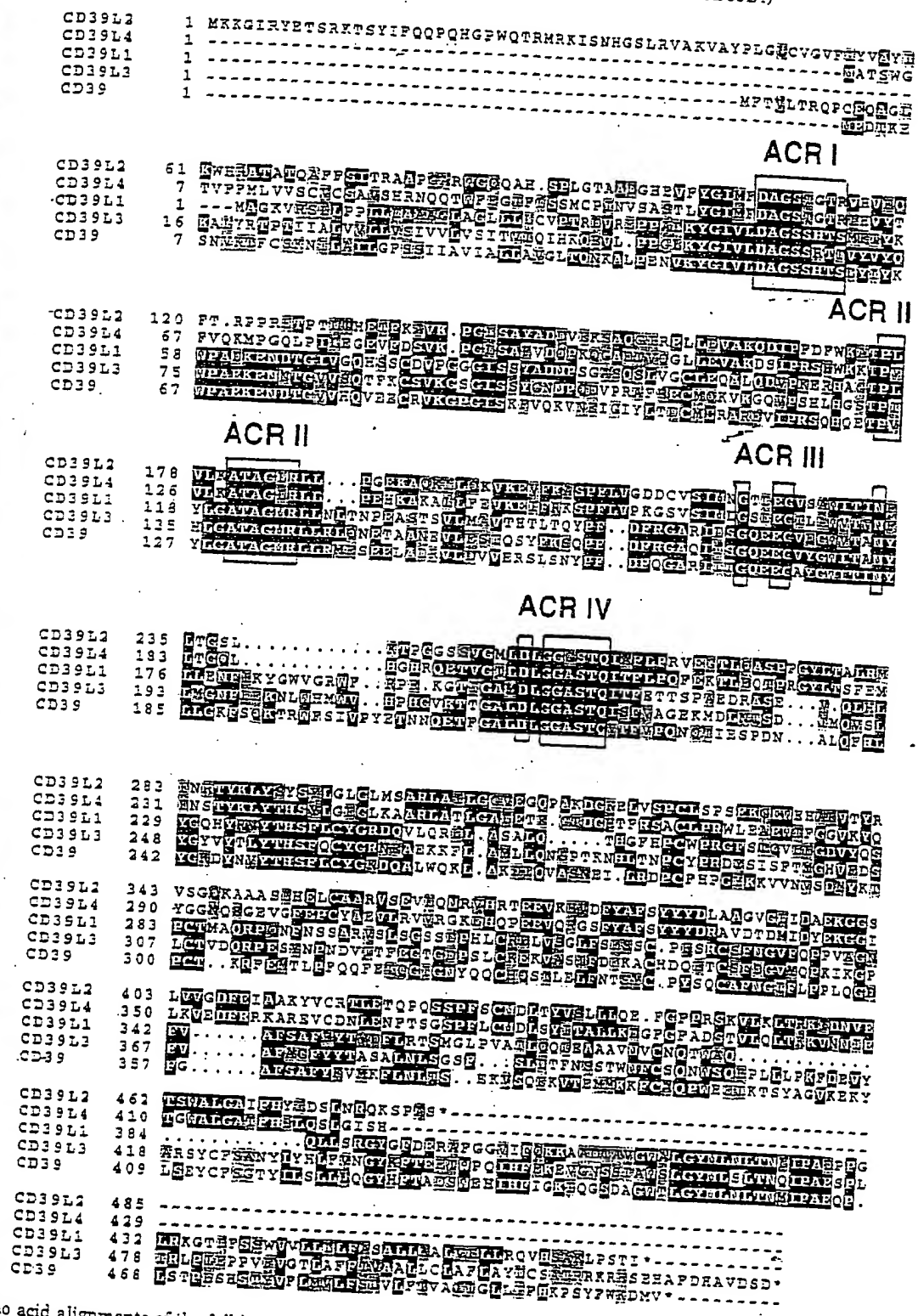
Northern analysis of members of the CD39-like gene family. cDNA clone inserts were removed by restriction digestion and separated by gel electrophoresis. Insert DNA was gel-purified and radiolabeled (Sambrook *et al.*, 1989). Radiolabeled cDNA was prehybridized at 65°C for 2 h with 20 μg of human Cot-1 DNA (GibcoBRL) and 100 μg of total human DNA (Sigma), before hybridization to Northern blots (Clontech, human multiple tissue Northern blots, Catalog No. 7760-1 and 7759-1) according to the manufacturer's instructions.

Mapping of CD39L2, CD39L3, and CD39L4. Members of the *CD39-Like* gene family were mapped in the human genome by PCR screening of the GeneBridge-4 radiation hybrid mapping panel obtained from the HGMP Resource Centre (Hinxton, UK) (Gyapay *et al.*, 1996). PCR-positive radiation hybrid clones were organized into the GeneBridge-4 HGMP-RC subset order using the HGMP radiation hybrid mapping World Wide Web (WWW) site (http://www.hgmp.mrc.ac.uk/cgi-bin/contig/rhmapper.pl), and mapping data for each gene were obtained from the Whitehead server (http://www.genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl). The chromosomal location for each gene was confirmed by PCR screening of the monochromosomal hybrids obtained from the HGMP Resource Centre. PCR primers were designed for the 3′ untranslated region (UTR) of each gene and titrated for a unique human-specific PCR product. *CD39L2*, Primer 1, 5′CTGCTTGAGTGACGTCTCTG3′; Primer 2, 5′CACATGAGGTTCAGCTCGTG3′; 94°C for 2 min; 38 cycles of 94°C for 20 s, 54°C for 20 s, 72°C for 20 s; 72°C for 2 min. Product size is 362 bp). *CD39L3*, Primer 1: 5′GTGAAGTGGCTGCCTTCAGG3′; Primer 2, 5′CCTTTGACTCGGGACTCCAG3′; 94°C for 2 min; 38 cycles of 94°C for 20 s, 56°C for 20 s, 72°C for 20 s; 72°C for 2 min. Product size is 281 bp). *CD39L4*. Primer 1, 5′GAACTGCTGCCTAACCACTC3′; Primer 2, 5′ATTGATGGGTCTTGGGATTGC3′; 94°C 2 for min; 38 cycles of 94°C for 20 s, 56°C for 20 s, 72°C for 20 s; 72°C for 2 min. Product size is 234 bp. PCR products were analyzed by electrophoresis through 3.5% NuSieve agarose gels (Flowgen).

RESULTS

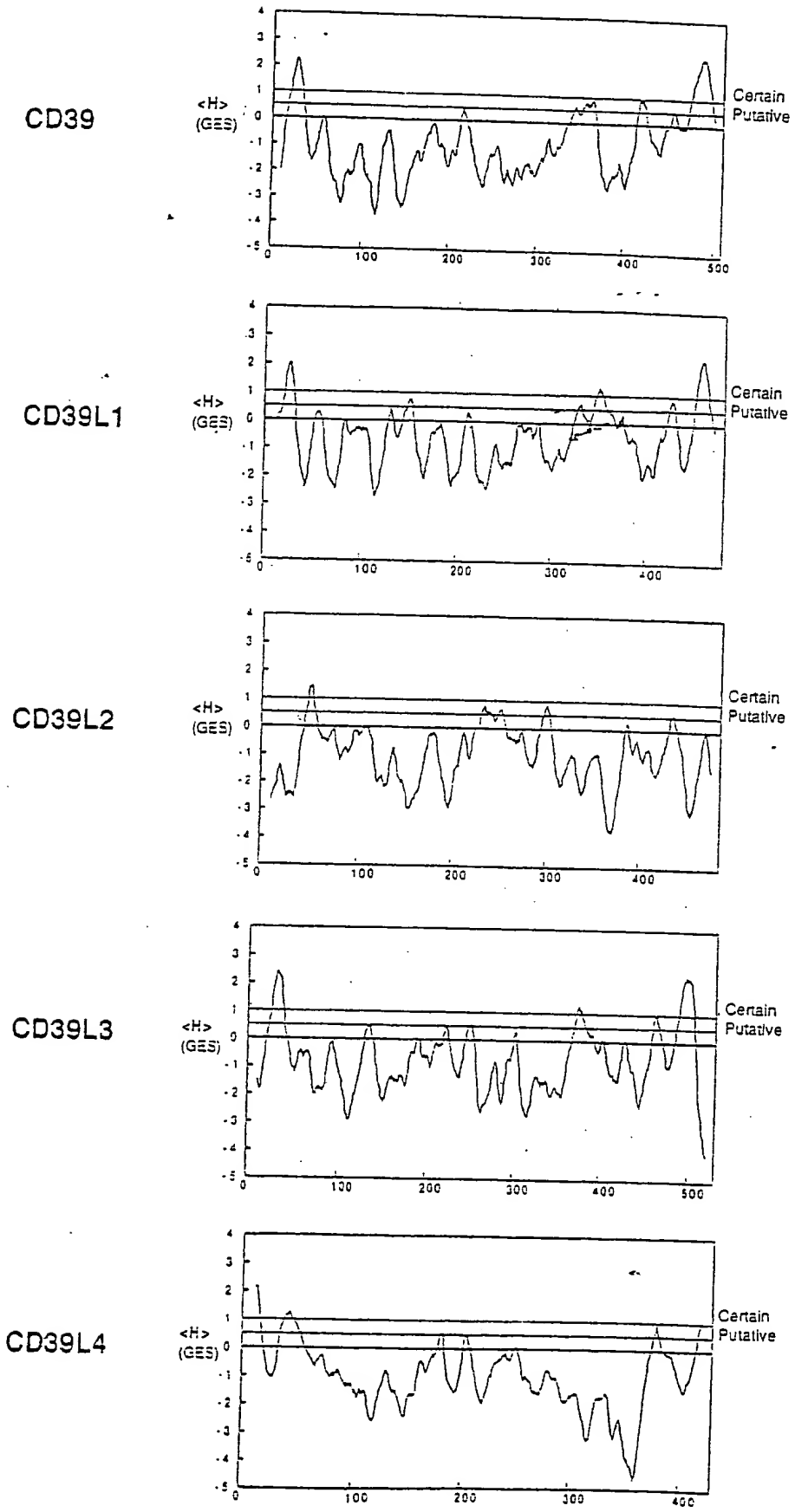
Isolation and Sequence Characterization of CD39L2

EST entries for human *CD39L2* were identified by a BLASTN search with the full-length *mNTPase* cDNA sequence (Chadwick *et al.*, 1998). A 1.7-kb ESTinsert (Accession No. H08436) was sequenced, and an ORF was identified that had high amino-acid homology to CD39 and CD39L1 and contained the ACRs II, III, and IV (data not shown). The insert of H08436 was used to isolate additional clones from a human adult breast epithelial cDNA library (ZR75), a human T-leukemia cell line J6 cDNA library (Jurkat), and a human keratinocyte stem cell cDNA library (KER). Of 23 cDNA clones that were isolated and sequenced, all but one (*CD39L2K8* isolated from the KER cDNA library: Accession No. AF039916) appeared to be alternatively spliced or unspliced. An ORF extending to nucleotide 1600 containing ACRs I–IV was identified within the 2762-bp cDNA sequence for *CD39L2K8*. Two ATG codons with a poor match to the consensus translation initiation site were found at nucleotide positions 148 and 232 [AUGUGAAUGA at 148 and ACAAGGAUGA at 232 versus consensus GCCGCCAUGG; (Kozak, 1989)]. Based on homology to a putative mouse homologue (see below), the ATG at nucleotide position 232 is the initiation codon. A single polyadenylation signal of AAUAAA was identified at nucleotide position 2700, 22 nucleotides 5′ of the poly(A) tail of *CD39L2K8*. The deduced amino acid sequence of CD39L2 is shown in Fig. 1. Hydrophobicity plots using Topred-II 1.1 (Claros and Von Heijne, 1994) predicted a single transmembrane segment at the N-terminal extreme of the protein, suggesting that CD39L2 has a short putative cytoplasmic tail and a large extracellular C-terminal domain (Fig. 2). There are two potential N-glycosylation sites in the predicted extracellular domain. A cAMP- and cGMP-dependent protein kinase and a protein kinase-C phosphorylation site are found directly after the initiation methionine codon (nucleotide 232).



Amino acid alignments of the full-length protein sequences for human members of the CD39-like gene family: CD39 (Accession ...13), CD39L1 (Accession No. U91510), CD39L2 (Accession No. AF039916), CD39L3 (Accession No. AF039917), and CD39L4 (Acces... AF039918). Identical residues are indicated by a black background, whereas conserved residues are indicated by a gray background. ...the sequences are indicated by a dot. The apyrase conserved regions I–IV are highlighted by the boxed sections. Alignments ...e using Pileup and Boxshade from the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin.

...tion and Sequence Characterization of CD39L3. ...was identified by a BLASTN search of the EST ...e at NCBI with the full cDNA sequence for hu... CD39 (Accession No. S73813). Three EST entries ...A clones from an endometrial tumor library were identified (Accession Nos. AA336644, AA338117, and AA337885), which led to the identification of an IMAGE library EST (Accession No. N72742) that was completely sequenced and had high amino acid homology to CD39 and CD39L1 (data not shown). The insert of N72742 was


CD39
CD39L1
CD39L2
CD39L3
CD39L4
<H>
(GES)
Certain
Putative
Amino acid position

used to screen the Jurkat, ZR75, and KER cDNA libraries, and a single cDNA clone was isolated from the KER library and sequenced (*CD39L3K1* Accession No. AF039917). The 1669-bp ORF in *CD39L3K1* contains ACRs I–IV, an ATG codon at position 83, and a single polyadenylation signal at position 2758. The deduced amino acid sequence of CD39L3 is shown in Fig. 1. Hydrophobicity plots as described above predict two potential transmembrane segments at the N- and C-terminal extremes of the protein (Fig. 2). There are seven potential extracellular N-glycosylation sites. A cAMP- and cGMP-dependent protein kinase site and a protein kinase-C phosphorylation site are located at the C-terminal extreme of the protein.

Isolation and Sequence Characterization of CD39L4

CD39L4 was identified by a TBLASTX search of the EST database at NCBI with the full cDNA sequence for the mouse *mNTPase* gene (Accession No. AF006482). A human EST clone was sequenced, and an ORF was identified extending to nucleotide 529 of 2260 nucleotides that contained ACR I only and an ATG codon at position 256. In the same reading frame, downstream of the stop codon at nucleotide 529, an ORF extending to nucleotide 1792 contained ACRs II, III, and IV. Further analysis of the nucleotide sequence revealed a putative intron with splice donor and acceptor sites that conform to the 5′ gt···3′ ag rule (Breathnach and Chambon, 1981) (splice donor CAGgtcacttatggagcctg at nucleotide position 470, acceptor ccatggacaaaatagGAC at position 710, exon sequence underlined). Removal of the 251-bp putative intron would result in a contiguous ORF containing ACRs I–IV. The assumption that this sequence does indeed constitute an intron was confirmed by sequencing three additional cDNA clones (*CD39L4J1–3*) from the Jurkat library, none of which contained the 251 bp. The cDNA sequence (Accession No. AF039918) contained a poly(A) tail, but no consensus polyadenylation sequence (Proudfoot, 1991). This is also the case for the mouse homologue (see below). The deduced amino acid sequence of CD39L4 is shown in Fig. 1. Hydrophobicity plots as described above predict a single transmembrane segment at the N-terminal extreme of the protein (Fig. 2). This is similar to the predicted topology of CD39L2 and different from that of CD39, CD39L1, and CD39L3. There are three potential extracellular N-glycosylation sites.

Expression of the Human Members of the CD39-like Family

Representative probes for each member of the *CD39*-like gene family were hybridized to human multiple tissue Northern blots (Fig. 3).

CD39. A 1.7-kb clone (Accession No. AA284456) corresponding to *CD39* was obtained from HGMP, and identity was confirmed by sequencing of the ends (Fig. 3). Two prominent signals are found at 3.2 kb (major) and 3.6 kb (minor). A strong 3.2-kb signal can be seen in adult spleen, peripheral blood leukocytes, small intestine, and placenta. An additional signal of approximately 9.5 kb, which could be a result of differential polyadenylation, can be seen in adult spleen and peripheral blood leukocytes. A weaker signal of approximately 1.8 kb in adult placenta may be a result of alternative splicing.

CD39L1. The expression of *CD39L1* has been described previously (Chadwick and Frischauf, 1997).

CD39L2. Hybridization of the *CD39L2K8* insert to the multiple tissue Northern blots (Fig. 3) resulted in two prominent signals of 2.6 kb (major) and 4.4 kb (minor) in all tissues studied. This is most likely due to differential polyadenylation.

CD39L3. A PCR product covering the coding sequence of *CD39L3* (courtesy of Tom Smith and Terry Kirley) was used for the Northern hybridizations (Fig. 3). A strong signal of approximately 3.0 kb can be seen in adult brain, pancreas, spleen, and prostate. Though moderate or low expression is seen in most other tissues, no signals was detected in adult liver and peripheral blood leukocytes. A weaker signal of approximately 1.8 kb found in adult pancreas may be the result of alternative splicing.

CD39L4. The insert of AA256016 was hybridized to the same Northern blots (Fig. 3), and a prominent signal of approximately 4.8 kb was seen in adult liver, kidney, prostate, testis, and colon. Considerably weaker expression was seen for all other tissues examined. Several smaller bands observed in tissues showing the strongest expression of *CD39L4* may be the result of differential polyadenylation or alternative splicing.

Mapping of Members of the CD39-like Family

The *CD39* gene has previously been mapped to human chromosome 10q23.1–q24.1 (Maliszewski *et al.*, 1994). The mouse homologue *cd39* has not been mapped, but on the basis of conserved gene order between the human and the mouse chromosomes, the likely map position for murine *cd39* would be on mouse chromosome 19 (DeBry and Seldin, 1996).

The *CD39L1* gene was previously mapped to human chromosome 9q34, whereas the mouse homologue *cd39L1* was mapped by FISH to mouse chromosome 2, the syntenic region of human 9q34 (Chadwick and Frischauf, 1997).

The *CD39L2* gene was mapped with a lod score of >19 to human chromosome 20 by PCR typing of the GeneBridge 4 Radiation Hybrid Mapping Panel (Gyapay *et al.*, 1996). *CD39L2* mapped 9.76 cR from D20S493 (typing data: 12012 02101 22000 00111 00110 01210 00110 01101 10121 00100 00120 11211

FIG. 2. Comparison of hydrophobicity predictions for the amino acid sequences of members of the human CD39-like gene family. Predictions were made using the Topred-II 1.1 program (Claros and von Heijne, 1994) (putative setting = 0.5; certain setting = 1.0).

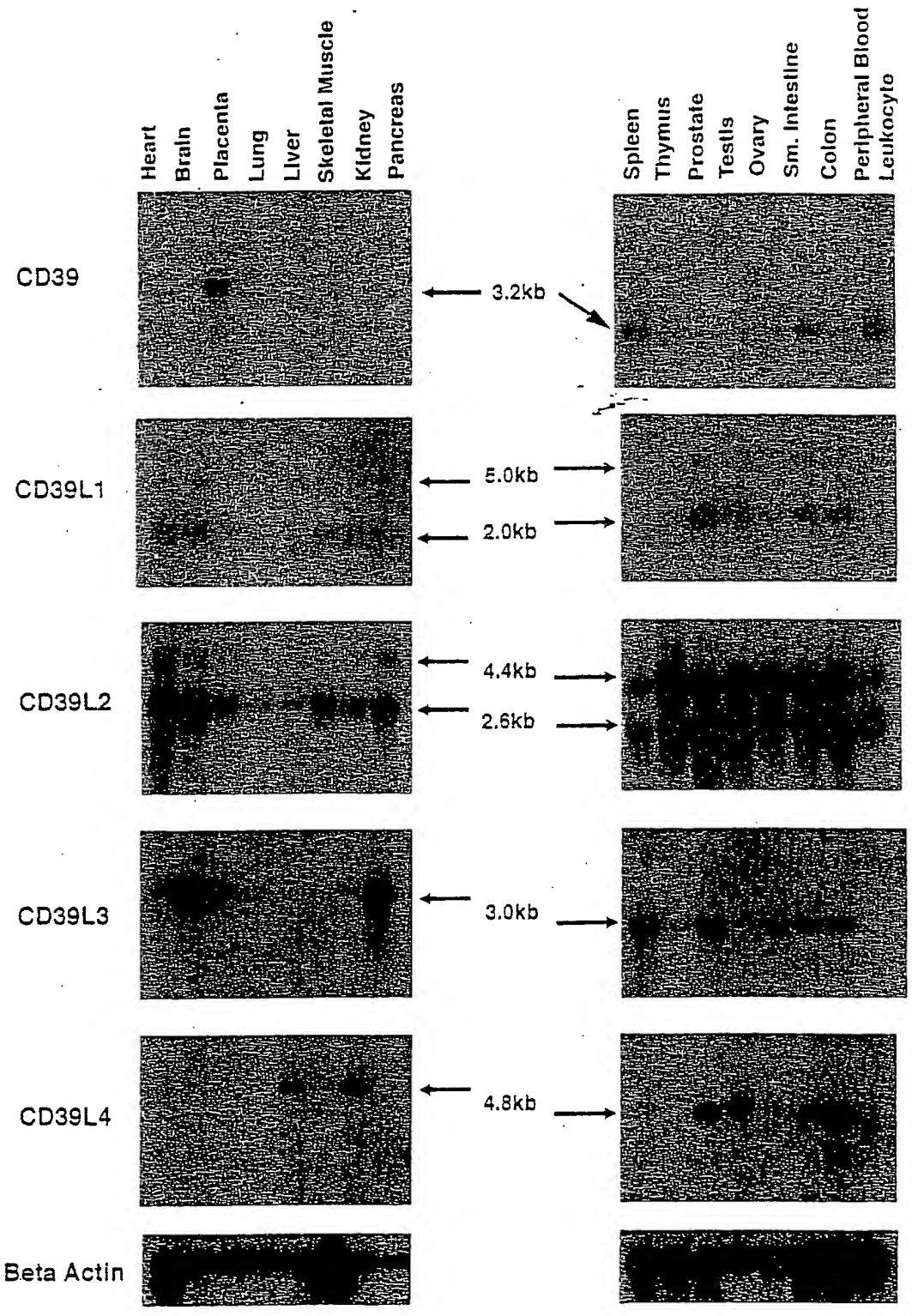


FIG. 3. Alignment of multiple tissue Northern hybridizations for members of the human CD39-like gene family.

00011 11012 01001 01102 00000 00000 001). Using the closest flanking markers (D20S184 and D20S99) also represented on the consensus map, this places *CD39L2* at chromosome band 20q11.2. The location of *CD39L2* on human chromosome 20 was confirmed by PCR analysis of monochromosomal mapping panels (Kelsell *et al.*, 1995). On the basis of synteny to human chromosome 20q11.2, the mouse homologue of *CD39L2* is expected to map to mouse chromosome 2 (DeBry and Seldin, 1996).

The *CD39L3* gene was mapped as described above to human chromosome 3, 5.76 cR from D3S3390 (data: 12002 02010 22000 00011 20000 00110 01001 00000 02022 11000 10001 00200 21100 00212 01010 10002

TABLE 1

Level of Nucleotide Sequence Identity between CD39L2 and CD39L3 with Mouse EST Entries

Mouse EST	Coverage[a]	Nucleotide identity (%)
AA611283	nt 232–580	92.5
AA647051	nt 740–994	86.7
AA638277	nt 1080–1600	82.0
AA271520	nt 207–379	79.0
W46136	nt 1550–1669	68.0

[a] Nucleotide (nt) position of human *CD39-like* gene.

1000 00011 001). Using the closest flanking markers described above (D3S1561 and D3S3564), this places *CD39L3* at chromosome band 3p21.3. The location of *CD39L3* on chromosome 3 was confirmed by PCR as *CD39L2*. On the basis of synteny, the mouse homologue of *CD39L3* is expected to map to mouse chromosome 9 (DeBry and Seldin, 1996).

The *CD39L4* gene was mapped as described above to human chromosome 14, 1.92 cR from D14S71 (data: 02102 02102 22000 01010 11021 01000 01010 10110 02121 21000 00010 00211 01001 10102 02012 00002 01111 01100 002). This places *CD39L4* at chromosome band 14q24. The chromosomal location of *CD39L4* was confirmed as described above. The mouse homologue of *CD39L4* (*mNTPase*, see below) was previously mapped to mouse chromosome 12 (Chadwick *et al.*, 1998).

Murine Homologues of the Human CD39-like Gene Family

The mouse homologues of the *CD39* and *CD39L1* genes have been described previously (Maliszewski *et al.*, 1994; Chadwick and Frischauf, 1997).

After the EST database at NCBI was searched by TBLASTN with the 2762-bp *CD39L2* cDNA sequence, three mouse ESTs showed significant DNA sequence identity (Table 1). The region of highest homology is within the coding region of *CD39L2*, with sequence divergence before the start codon and after the stop codon. An alignment of the DNA sequence for EST AA611283 against *CD39L2* shows decreasing homology and several gaps upstream of nucleotide 232 (data not shown), which corresponds to the second potential initiation methionine for *CD39L2*. Downstream of nucleotide 232 there are no gaps in an alignment with greater than 90% sequence identity. It is very likely that these mouse ESTs represent the mouse *cd39L2* gene.

Several mouse EST entries with significant sequence identity to human *CD39L3* were identified in a BLASTN search using the *CD39L3* cDNA sequence. Two representative mouse ESTs are listed in Table 1. As with *CD39L2*, the coding region is highly homologous to *CD39L3*, indicating that these mouse ESTs represent the mouse *cd39L3* gene.

Human *CD39L4* was initially identified using the *mNTPase* sequence. Alignment of the full sequence of *CD39L4* against *mNTPase* shows 88% sequence identity between nucleotides 247 and 1533, the start and stop codons of *CD39L4* (data not shown). This makes it very likely that *mNTPase* is the mouse homologue of *CD39L4*, a conclusion supported by the map position of *mNTPase* in a region of synteny to the chromosomal location of human *CD39L4* (see above).

Identification of a Drosophila Gene with High Homology to CD39L2 and CD39L4

A *D. melanogaster CD39-like* gene was identified by a TBLASTX search of the EST database using the human *CD39L2* cDNA sequence. Five *Drosophila* EST entries were identified (Accession No. AA391625, AA390461, AA201196, AA246996, and AA567512), and a consensus sequence was generated and used for a BLASTN search against EMBL/GenBank entries. A single *D. melanogaster* genomic entry (Accession No. AC002032) was identified showing 100% sequence identity to three regions of the EST consensus sequence. Alignment of the EST consensus against the genomic sequence identified three exons that conform to the 5′ gt···3′ ag rule (Breathnach and Chambon, 1981). Exon 4 was identified on the basis of reading frame homology to the CD39L2 and CD39L4 proteins. An ATG codon was identified in exon 1, a stop codon in exon 4. Table 2 summarizes the genomic structure of *dmNTPase*. The predicted amino acid sequence of the *D. melanogaster CD39-like* gene containing the ACRs I–IV is shown in Fig. 4, aligned against the gene family members with the highest homology. Three N-glycosylation consensus sites are found in the putative extracellular domain, and two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites are found in the putative N-terminal cytoplasmic domain. Hydrophobicity plots as described above predict a single transmembrane segment at the N-terminal extreme of the dmNTPase protein. The topology of dmNTPase is therefore most similar to the predicted topology of the CD39L2 and CD39L4 proteins.

DISCUSSION

This paper reports the cloning, mapping, and expression of three new human members of the *CD39-like* gene family. The identification of the new genes made extensive use of gene and EST databases. Searching databases has replaced low-stringency cDNA library screening and has identified related genes that show significant amino acid homology, but are unlikely to cross-hybridize at the nucleotide level. Figure 1 shows an alignment of the five human members of the gene family. A high level of amino acid identity can be seen for all five proteins, and the presence of ACRs I–IV suggests that they are likely to have apyrase and/or ATPase activity. The N-terminus is most highly conserved, consistent with the idea that it contains the active site. It is clear from the protein alignments in Fig. 1 that many other residues in addition to the ACRs

TABLE 2

The Genomic Structure of the *Drosophila* NTPase Gene

Exon	Acceptor	Donor	Exon size	Intron size
1	—	CTCgtagacccgatgaacttggc	>193 bp	85 bp
2	tttgtcttttcgccttttttcagCTG	TTGgtgtgtaccatttgtacttat	141 bp	79 bp
3	aacgactgacaacgatgcagGCT	AAGgtgagtcccggcgcacttaa	1129 bp	60 bp
4	actcgttttaaattggtgcagCTT	—	>112 bp	—

Note. The intronic sequence is given in lowercase, with the conforming 5′ gt and 3′ ag sequences in boldface. Exon sequence is given in uppercase letters. Approximate total coverage is 1.9 kb.

are conserved within the N-terminal region of the proteins. This is true not only of the human members of the gene family; alignment against members from nematode, yeast, and plants also shows conserved residues in addition to the ACRs (data not shown).

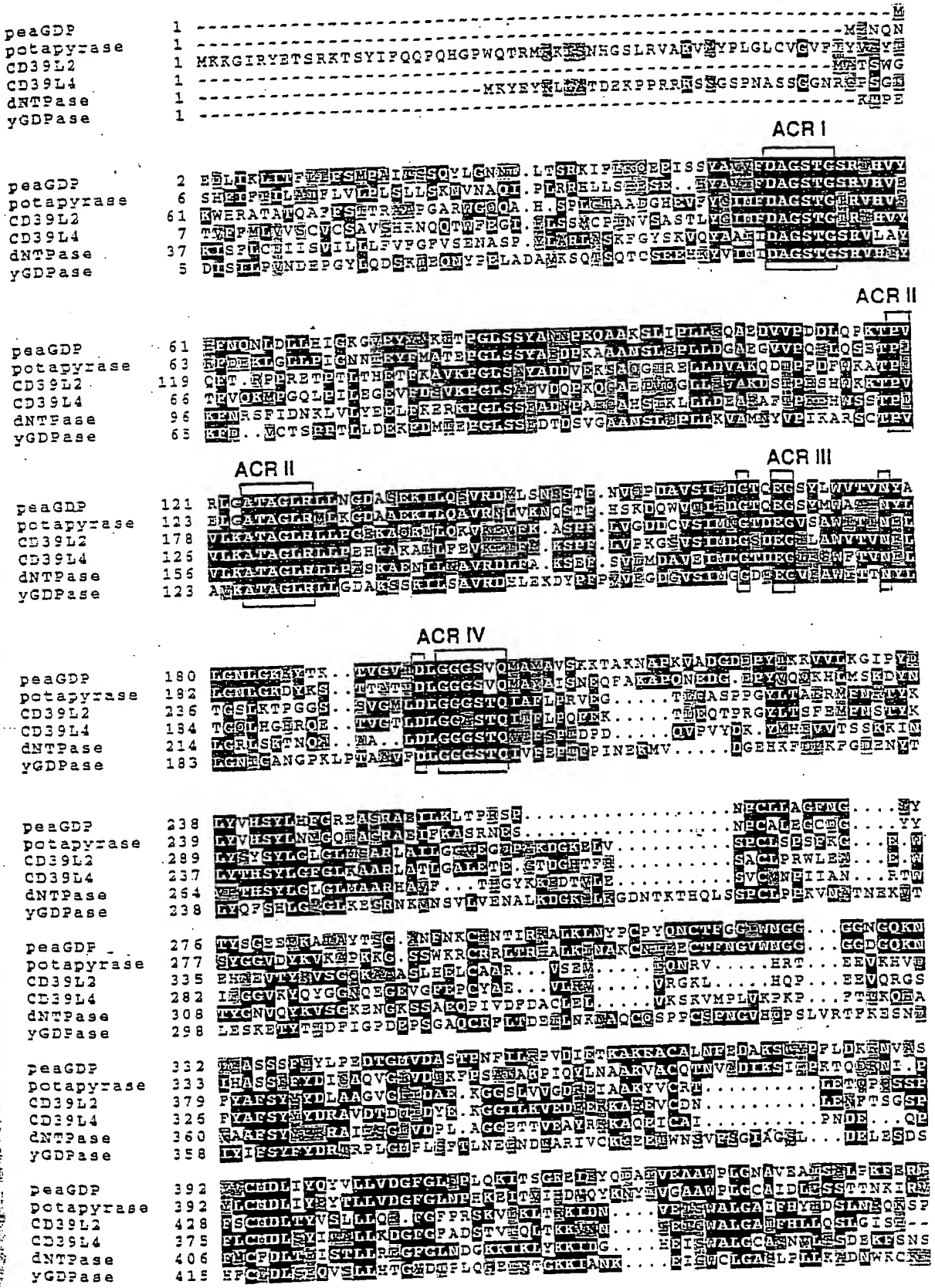
Topology predictions for the deduced amino acid sequence of each member of the family are shown in Fig. 2. The CD39, CD39L1, and CD39L3 proteins have very similar hydrophobicity profiles. As predicted with other vertebrate members of the family, this suggests that the proteins have short cytoplasmic N-terminal and C-terminal domains, with a large extracellular domain (Maliszewski *et al.*, 1994; Chadwick and Frischauf, 1997; Kirley, 1997). The human CD39L2 and CD39L4 proteins also show a potential transmembrane domain at the N-terminus of the protein, but do not appear to have a second transmembrane domain. This would suggest that CD39L2 and CD39L4 have a short cytoplasmic N-terminus and a large extracellular C-terminus. This is very similar to the predicted membrane topology of the NTPases described in yeast (Abeijon *et al.*, 1993), potato (Handa and Guidotti, 1996), garden pea (Lin, 1989), and *D. melanogaster* (data not shown). The NTP1 gene from *T. gondii* (Bermudes *et al.*, 1994) and potato apyrase (Handa and Guidotti, 1996) and an NTPase from mosquito saliva (Champagne *et al.*, 1995) are all soluble apyrases. It is therefore possible that the predicted first transmembrane domain in CD39L2 and CD39L4 is also a signal sequence that is not present in the mature protein.

The lymphoid cell activation antigen CD39 is known to be heavily N-glycosylated (Kansas *et al.*, 1991). The predicted protein sequences of CD39L1–4 all contain several putative N-linked glycosylation consensus sequences, suggesting that each member of the family is a glycoprotein. Numerous cysteine residues are found in the putative extracellular domain of each protein. Due to the oxidizing extracellular media, these may form disulfide bonds to stabilize, fold, or dimerize the proteins. Consensus sequences for protein kinase C and cAMP- and cGMP-dependent protein kinase phosphorylation can be found in the cytoplasmic domains of all but CD39L4 of the human proteins. This suggests that the proteins may be regulated by cytoplasmic protein kinases.

Close examination of Fig. 1 shows that CD39L2 and CD39L4 show higher amino acid homology to one another than to CD39, CD39L1, and CD39L3. The predicted protein sequences of CD39L2 and CD39L4 are not only considerably shorter at the C terminus, but also diverge significantly after ACRIV from CD39, CD39L1, and CD39L3. We have therefore grouped together the following genes as members of class I: *CD39*, *cd39*, rat *cd39*, *CD39L1*, *cd39l1*, *CD39L3*, *cd39l3*, and the chicken muscle ecto-ATPase. The following genes belong to class II: *CD39L2*, *cd39l2*, *CD39L4*, and *cd39l4*. From the alignment of CD39L2 and CD39L4 with the NTPases from yeast, garden pea, potato, *D. melanogaster* (Fig. 4) and *T. gondii NTP1* (not shown), it is clear that these two proteins are more closely related to the plant and invertebrate proteins than to CD39, CD39L1, and CD39L3. Class II members could be the soluble vertebrate ectonucleotidases postulated by Lewis-Carl and Kirley (1997). Such enzymes might have important clinical applications in anti-thrombotic agents (Marcus *et al.*, 1997), in the regulation of neurotransmission by ATP (Todorov *et al.*, 1997), and as anti-tissue graft rejection agents (Candinas *et al.*, 1996; Robson *et al.*, 1996). It is well established that ATP acts as a rapid neurotransmitter in smooth muscle, peripheral ganglia, and the brain (Zimmermann, 1994; Vizi *et al.*, 1997). ATP binds to the P_2X purinoreceptor, a ligand-gated ion channel found in neurons and smooth muscle (Chen *et al.*, 1995). It has recently been shown in vertebrates that the release of ATP by stimulated sympathetic neurons is accompanied by the neuronal release of soluble nucleotidases (Todorov *et al.*, 1997). Human *CD39L2* shows high levels of expression in adult brain (Fig. 3). It is therefore possible that *CD39L2* may be the soluble nucleotidase described. The isolation of a related gene from *D. melanogaster* may provide a manipulable system to evaluate further the role of ecto-nucleotidases during the development and maintenance of neurons and also to study ecto-nucleotidases in nonneuronal tissues.

An interesting genetic association between divalent

FIG. 4. Alignment of the *Drosophila melanogaster* CD39-like predicted protein sequence (dNTPase; Accession No. AF041048) to the most closely related members of the CD39-like gene family; peaGDP, garden pea NTPase (Accession No. P52194); potapyrase, potato RROP1 gene (Accession No. gi|1381633); CD39L2, human CD39L2 (Accession No. AF039916); CD39L4, human CD39L4 (Accession No. AF039915); and yGDPase, yeast yGDA1 gene (Accession No. sp|P32621).


ACR I
ACR II
ACR II
ACR III
ACR IV
peaGDP
potapyrase
CD39L2
CD39L4
dNTPase
yGDPase

cation-dependent ATPase activity and audiogenic seizure susceptibility (AGS) in mice is well documented (Palayoor and Seyfried, 1984; Allen and Seyfried, 1994; Banko *et al.*, 1997). Mice susceptible to audiogenic seizure show generalized convulsions when subjected to loud high-frequency sound. Three genetic loci in mice have been associated with AGS on mouse chromosomes 4, 7, and 12 (Neumann and Seyfried, 1990; Neumann and Collins, 1991). The mouse homologue of *CD39L4* has been mapped to mouse chromosome 12 (Chadwick *et al.*, 1998). The locus for AGS on mouse chromosome 12 (Asp-1) is linked to the genetic marker D12Nyu1, 30.2 cM from the centromere of chromosome 12 (Seyfried and Glaser, 1981). Mouse *cd39l4* is linked to D12Mit4, 28.4 cM from the centromere of chromosome 12. It is therefore possible that *cd39l4* is Asp-1 and responsible for AGS in mice. NTPase activity in mouse brain has also been associated with a locus on mouse chromosome 2 (Allen and Seyfried, 1994). The map location of the human members of the *CD39-like* gene family allows a putative map location of the murine homologues on the basis of conserved gene order between the human and the mouse chromosomes (DeBry and Seldin, 1996). *cd39l1* and *cd39l12* map to mouse chromosome 2, and both are expressed in human brain and are therefore likely to be expressed in mouse brain (Fig. 3). The murine homologue of *CD39L2* potentially maps to distal mouse chromosome 2 (DeBry and Seldin, 1996). The mouse *cd39l1* gene has been mapped to proximal mouse chromosome 2 (Chadwick and Frischauf, 1997). The NTPase activity in mouse brain is also linked to proximal mouse chromosome 2 (Allen and Seyfried, 1994); therefore mouse *cd39l1* is a candidate for the NTPase activity observed in mouse brain.

Examination of the expression patterns of each of the five members of the *CD39-like* gene family (Fig. 3) shows expression of each gene in several different tissues. All tissues examined showed expression of at least one member of the gene family, while some tissues, such as prostate, show expression of all five members. Biochemical evidence of at least two distinct ecto-ATPases has been presented for liver (Yamaguchi and Ohnishi, 1977; Karasaki *et al.*, 1980), which is confirmed by the Northern analysis showing the presence of *CD39, CD39L2,* and *CD39L4* in human adult liver.

The results presented here show that there are many more NTPases than previously thought. There may be additional variations even without further gene discoveries, since in Northern analysis of the five human members of the gene family shown in Fig. 3, several hybridizing bands were observed, indicating the presence of alternatively spliced mRNAs. This was confirmed during the isolation of cDNA, when several different cDNA clones were found for each gene (data not shown). These included some mRNA species that appeared to contain introns and were therefore unspliced, while others showed splicing of exons that would lead to truncated proteins as was observed with a cDNA clone for the *mcd39l1* gene (Chadwick and Frischauf, 1997). Some cDNA clones showed the lack of one or more of the ACRs. It is not clear whether these alternatively spliced products are translated and if so whether they show any activity. If they do result in the production of functional protein, this would increase the variety and complexity of NTPases present in cells even further.

ACKNOWLEDGMENTS

We thank the HGMP Resource Centre (Hinxton, UK) for the rapid and efficient provision of resources (GeneBridge 4 radiation hybrid mapping panels, monochromosomal hybrid mapping panels, and the IMAGE cDNA clones). We are very grateful to Tom Smith and Terry Kirley for the provision of a full-length *CD39L3* PCR product for the Northern analysis. We are also grateful to Phil Jones and Fiona Watt for the human keratinocyte stem cell cDNA library, Rodger White and Malcolm Parker for the ZR75 Lambda ZAP cDNA library, Jenny Dunne and Mike Owen for the Jurkat cell line cDNA library, and Ian Goldsmith and colleagues for oligonucleotide synthesis. This work was supported by the Imperial Cancer Research Fund.

REFERENCES

Abeijon, C., Yanagisawa, K., Mandon, E. C., Hausler, A., Moremen, K., Hirschberg, C. B., and Robbins, P. W. (1993). Guanosine diphosphatase is required for protein and sphingolipid glycosylation in the Golgi lumen of *Saccharomyces cerevisiae*. *J. Cell Biol.* 122: 307–323.

Allen, K. M., and Seyfried, T. N. (1994). Genetic analysis of nucleotide triphosphatase activity in the mouse brain. *Genetics* 137: 257–265.

Banko, M. L., Allen, K. M., Dolina, S., Neumann, P. E., and Seyfried, T. N. (1997). Genomic imprinting and audiogenic seizures in mice. *Behav. Genet.* 27: 465–475.

Bermudes, D., Peck, K. R., Afifi, M. A., Beckers, C. J., and Joiner, K. A. (1994). Tandemly repeated genes encode nucleoside triphosphate hydrolase isoforms secreted into the parasitophorous vacuole of *Toxoplasma gondii*. *J. Biol. Chem.* 269: 29252–29260.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50: 349–383.

Candinas, D., Koyamada, N., Miyatake, T., Siegel, J., Hancock, W. W., Bach, F. H., and Robson, S. C. (1996). Loss of rat glomerular ATP diphosphohydrolase activity during reperfusion injury is associated with oxidative stress reactions. *Thromb. Haemost.* 76: 807–812.

Chadwick, B. P., and Frischauf, A. M. (1997). Cloning and mapping of a human and mouse gene with homology to ecto-ATPase genes. *Mamm. Genome* 8: 668–672.

Chadwick, B. P., Williamson, J., Sheer, D., and Frischauf, A.-M. (1998). cDNA cloning and chromosomal mapping of a mouse gene with homology to NTPases. *Mamm. Genome* 9: 162–164.

Champagne, D. E., Smartt, C. T., Ribeiro, J. M., and James, A. A. (1995). The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5′-nucleotidase family. *Proc. Natl. Acad. Sci. USA* 92: 694–698.

Chen, W., Wieraszko, A., Hogan, M. V., Yang, H. A., Kornecki, E., and Ehrlich, Y. H. (1996). Surface protein phosphorylation by ecto-protein kinase is required for the maintenance of hippocampal long-term potentiation. *Proc. Natl. Acad. Sci. USA* 93: 8685–8693.

Chen, Z. P., Levy, A., and Lightman, S. L. (1995). Nucleotides as extracellular signalling molecules. *J. Neuroendocrinol.* 7: 83–96.

Claros, M. G., and von Heijne, G. (1994). TopPred II: An improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* 10: 685–686.

DeBry, R. W., and Seldin, M. F. (1996). Human/mouse homology relationships. *Genomics* 33: 337–351.

Di Virgilio, F. (1995). The P2Z purinoceptor: An intriguing role in immunity, inflammation and cell death. *Immunol. Today* 16: 524–528.

Dunne, J., Hanby, A. M., Poulsom, R., Jones, T. A., Sheer, D., Chin, W. G., Da, S. M., Zhao, Q., Beverley, P. C., and Owen, M. J. (1995). Molecular cloning and tissue expression of FAT, the human homologue of the *Drosophila* fat gene that is located on chromosome 4q34–q35 and encodes a putative adhesion molecule. *Genomics* 30: 207–223.

Felts, P. A., and Smith, K. J. (1996). Changes in the distribution of a calcium-dependent ATPase during demyelination and remyelination in the central nervous system. *J. Neurocytol.* 25: 171–180.

Filippini, A., Taffs, R. E., Agui, T., and Sitkovsky, M. V. (1990). Ecto-ATPase activity in cytolytic T-lymphocytes: Protection from the cytolytic effects of extracellular ATP. *J. Biol. Chem.* 265: 334–340.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J. F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J., and Goodfellow, P. N. (1996). A radiation hybrid map of the human genome. *Hum. Mol. Genet.* 5: 339–346.

Handa, M., and Guidotti, G. (1996). Purification and cloning of a soluble ATP-diphosphohydrolase (apyrase) from potato tubers (*Solanum tuberosum*). *Biochem. Biophys. Res. Commun.* 218: 916–923.

Kansas, G. S., Wood, G. S., and Tedder, T. F. (1991). Expression, distribution, and biochemistry of human CD39. Role in activation-associated homotypic adhesion of lymphocytes. *J. Immunol.* 146: 2235–2244.

Karasaki, S., Suh, M. H., Salas, M., and Raymond, J. (1980). Cell surface adenosine 5'-triphosphatase as an in vitro marker of the lineage and cytodifferentiation of oncogenic epithelial cells from rat liver parenchyma. *Cancer Res.* 40: 1318–1328.

Kegel, B., Braun, N., Heine, P., Maliszewski, C. R., and Zimmermann, H. (1997). An ecto-ATPase and an ecto-ATP diphosphohydrolase are expressed in rat brain. *Neuropharmacology* 36: 1189–1200.

Kelsell, D. P., Rooke, L., Warne, D., Bouzyk, M., Cullin, L., Cox, S., West, L., Povey, S., and Spurr, N. K. (1995). Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. *Ann. Hum. Genet.* 59: 233–241.

Kirley, T. L. (1997). Complementary DNA cloning and sequencing of the chicken muscle ecto-ATPase. Homology with the lymphoid cell activation antigen CD39. *J. Biol. Chem.* 272: 1076–1081.

Kozak, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* 108: 229–241.

Lewis-Carl, S., and Kirley, T. L. (1997). Immunolocalization of the ecto-ATPase and ecto-apyrase in chicken gizzard and stomach. Purification and N-terminal sequence of the stomach ecto-apyrase. *J. Biol. Chem.* 272: 23645–23652.

Lin, S. H. (1989). Localization of the ecto-ATPase (ecto-nucleotidase) in the rat hepatocyte plasma membrane. Implications for the functions of the ecto-ATPase. *J. Biol. Chem.* 264: 14403–14407.

Maliszewski, C. R., Delespesse, G. J., Schoenborn, M. A., Armitage, R. J., Fanslow, W. C., Nakajima, T., Baker, E., Sutherland, G. R., Poindexter, K., Birks, C., *et al.* (1994). The CD39 lymphoid cell activation antigen. Molecular cloning and structural characterization. *J. Immunol.* 153: 3574–3583.

Marcus, A. J., Broekman, M. J., Drosopoulos, J. H., Islam, N., Alyonycheva, T. N., Safier, L. B., Hajjar, K. A., Posnett, D. N., Schoenborn, M. A., Schooley, K. A., Gayle, R. B., and Maliszewski, C. R. (1997). The endothelial cell ecto-ADPase responsible for inhibition of platelet function is CD39. *J. Clin. Invest.* 99: 1351–1360.

Neumann, P. E., and Collins, R. L. (1991). Genetic dissection of susceptibility to audiogenic seizures in inbred mice. *Proc. Natl. Acad. Sci. USA* 88: 5408–5412.

Neumann, P. E., and Seyfried, T. N. (1990). Mapping of two genes that influence susceptibility to audiogenic seizures in crosses of C57BL/6J and DBA/2J mice. *Behav. Genet.* 20: 307–323.

Palayoor, S. T., and Seyfried, T. N. (1984). Genetic association between Ca^{2+}-ATPase activity and audiogenic seizures in mice. *J. Neurochem.* 42: 1771–1774.

Plesner, L. (1995). Ecto-ATPases: Identities and functions. *Int. Rev. Cytol.* 158: 141–214.

Proudfoot, N. J. (1991). High altitude termination. Regulation of transcription elongation and termination: A Keystone Symposium, Keystone, CO, April 19–25, 1991. *New Biol.* 3: 851–854.

Robson, S. C., Candinas, D., Winkler, H., Kopp, C., Hancock, W. W., and Bach, F. H. (1996). Immunobiology of discordant xenograft rejection and potential therapeutic modalities. *Transplant Proc.* 28: 1154–1155.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Seyfried, T. N., and Glaser, G. H. (1981). Genetic linkage between the AH locus and a major gene that inhibits susceptibility to audiogenic seizures in mice. *Genetics* 99: 117–126.

Todorov, L. D., Mihaylova-Todorova, S., Westfall, T. D., Sneddon, P., Kennedy, C., Bjur, R. A., and Westfall, D. P. (1997). Neuronal release of soluble nucleotidases and their role in neurotransmitter inactivation. *Nature* 387: 76–79.

Vizi, E. S., Liang, S. D., Sperlagh, B., Kittel, A., and Juranyi, Z. (1997). Studies on the release and extracellular metabolism of endogenous ATP in rat superior cervical ganglion: Support for neurotransmitter role of ATP. *Neuroscience* 79: 893–903.

Walter, J., Schnolzer, M., Pyerin, W., Kinzel, V., and Kubler, D. (1996). Induced release of cell surface protein kinase yields CK1- and CK2-like enzymes in tandem. *J. Biol. Chem.* 271: 111–119.

Wang, T. F., and Guidotti, G. (1996). CD39 is an ecto-(Ca^{2+}, Mg^{2+})-apyrase. *J. Biol. Chem.* 271: 9898–9901.

Yamaguchi, K., and Ohnishi, T. (1977). Surface ATPase activity at cell–cell contacts in hepatic parenchymal cells and in cAMP-treated hepatoma cells in monolayer culture. *Histochemistry* 54: 191–199.

Zimmermann, H. (1994). Signalling via ATP in the nervous system. *Trends Neurosci.* 17: 420–426.

Zimmermann, H. (1996). Biochemistry, localization and functional roles of ecto-nucleotidases in the nervous system. *Prog. Neurobiol.* 49: 589–618.